Application of Trajectory Data Clustering in CRM : A Case Study

V. TANUJA and P. GOVINDARAJULU

Department of Computer Science, Department of Computer Science
S.V. University, Tirupati, AP, INDIA, S.V. University , Tirupati, AP , INDIA

Summary
The role of data mining in today's business scenario is significant. Business entities, scientific research, social life, and many more real world situations rely on data processing and decision making based on mining knowledge masses of data available. Most of the real time data are trajectory centric. Trajectory data mining provide a better decision making. Trajectory data clustering in the current trend is in progress in trajectory data mining. As the present business picture is customer driven, the part of customer relationship management (CRM) is expected and invited. In this paper efforts are made to present the discussion on trajectory data clustering. A new algorithm is proposed for trajectory data clustering. Trajectory data relevant to public transportation is adopted processed, and the results are applied to suggest improvements in CRM with respect to the organization from where the data is collected.

Key words:
CRM, data mining, mobile

1. Introduction

1.1 Introduction to Trajectory Data Mining

The increasing pervasiveness of location acquisition technologies has enabled collection of very large trajectory datasets for different types of moving objects [16]. Discovered useful patterns from the movement behavior of moving objects are very valuable knowledge and useful to variety of real time applications. Considerable efforts have been applied to discover trajectory patterns of moving objects by using variety of data mining techniques [16]. Location acquisition and mobile computing technologies are generating very large spatial trajectory data [17]. Discovery of trajectory patterns is very useful in learning interactions between moving objects. Tree data structures are used for finding similarity between two trajectories. An identical sub-tree represents a similar sub structure between trees, whereas disjoint mapped nodes indicate no similar structure between the two trees [1]. Varieties of trajectory patterns have been proposed in the literature of trajectories. A trajectory of a moving object is a discrete trace that the moving object travels in geographical space [16]. The huge volume of spatial trajectories enables opportunities for analyzing the mobility patterns of moving objects, which can be represented by an individual trajectory containing a certain pattern or a group of trajectories sharing similar patterns [17]. Efficient storage of very large trajectory databases is the fundamental problem in trajectory management [21].

Mobility of transportation vehicles: A large number of GPS-equipped vehicles (such as taxis, buses, vessels, and aircrafts) have appeared in our daily life. For instance, many taxis in major cities have been equipped with a GPS sensor, which enables them to report a time-stamped location with a certain frequency [17]. Trajectories of moving objects are useful in finding knowledge such as moving patterns, moving Group patterns, finding location of a specific object or service etc. Zaiben Chen et al. [23] Proposed an algorithm to find the k Best-Connected Trajectories (k-BCT) from a database such that the k-BCT best connect the designated locations geographically.

Wireless networks and GPS are the two main sources of trajectory data for moving objects. The technologies like GPS provide a considerably more precise positioning. The behavior of the mobility of people changes over time. For instance, a new place of work, opening and closure of shops or changed means of transportation normally influence the mobile behavior. It is therefore important that algorithms can easily incorporate structural changes and adapt to new patterns in mobile behavior. Due to physical and faster an object moves, the more frequently an object's technical limitations during data collection and storage, uncertainty will arise which is an inherent characteristic of spatiotemporal data. While it can be broadly assumed that time is delivered with high accuracy, uncertainty of location varies with the applied technology between a few meters (GPS) and kilometers (GSM). In addition, the accuracy is greatly influenced by the sampling rate. The location is to be reported to sustain a given level of spatial uncertainty. Background knowledge as well as certain assumptions about movement behavior helps to reduce the uncertainty in data [15].

People have been recording their real-world movements in the form of spatial trajectories, passively and actively, for a long time [22]. One branch of trajectory analysis is meant for measuring similarity between trajectories. Euclidean distance and Dynamic Time Wrapping are two methods for finding similarity measure between two trajectories and these methods are failed in finding similarity of trajectories because they require multiple parameters to be set. In

Manuscript received January 5, 2017
Manuscript revised January 20, 2017

general, a trajectory is defined as a sequence of multi-dimensional points. A stay point is an important point in the trajectory with respect to processing or servicing where the user or object stayed for a while.

ZHENG Y et al. [25] proposed a location-history-based recommended system, which estimates the similarity between users in terms of their movements in geographical spaces. State-of the art computing technologies and location finding services have greatly supported human mobility. OznurKirmemisAlkan et al. [12] stated that, frequent pattern mining is a fundamental research topic that has been applied to different kinds of databases and it has been studied extensively by data mining researchers. In the trajectory literature many methods have been proposed for obtaining only a specific type of trajectory patterns and nowadays trajectory data is available abundantly [24].

- Trajectory preprocessing step is the most important step in all the trajectory data mining tasks and some of the preprocessing techniques are – noise filtering, segmentation, and map matching etc.
- Finding sequential patterns from a single trajectory or multiple trajectories is one of the dynamic research area in trajectory data management.

The main goal of the trajectory pattern data mining is to find hot regions from the trajectories, and then find sequential relationships among the hot regions, and then use these results in many real time applications such as clustering, classification, association and many other applications. There exist many prediction techniques that use vector based, pattern based, and association based models in order to predict location of users at the specific time given. Some authors have used association rules for storing movement behaviors of users and at the same time these association rules are represented in the trajectory pattern tree. Signature tree is also used by some authors as an indexing structure for deriving and managing sequential pattern relationship.

1.2 Introduction to Customer Relationship Management:

The concept of Customer Relationship Management has its ancestries in relationship marketing, which brought new approach to relations with customers, creating at the same time new market norms. CRM is a business strategy focused on maximizing stakeholder value through winning, growing, and retaining the right customers. We can distinguish here two important elements. CRM requires collection, storage and analysis of customer centric data to go for decision making [2]. First of all, concentrating on the most important from company's perspective customers and retaining long-term relationship with them. That is why it is essential to collect consequently customers' opinions, complaints and new needs. In this way it is possible to approach a client more individually, and make them feel important for the company, because each company is worth as much as customer values it [10].

CRM is recommended for establishing exceptional relationships with customers and for adding more value to goods and services than what is possible through traditional transaction practices. Traditional marketing was focused in gaining customers. Today it is time to retain the customers. The new CRM paradigm reflects a change in the traditional marketing. Customer retention is essential through great service, trust and, relationship. Then relationship marketing is not only about the 4Ps ((product, price, place, and promotion) but also long-term relationships with people with pace. An early definition of relationship marketing is provided by Gronroos, ‘‘The role of relationship marketing is to identify, establish, maintain and enhance relationships with customers and other stakeholders, at a profit, so that the objectives of all other parties involved are met; and that this is done by a mutual exchange and fulfillment of promises’’.

Within the present business scenario, characterized by an increasingly aggressive competence, the fight to win customers is stronger every day. Companies that enter to compete in a new market weaken the already existing and solid ones, due to the new methods of doing and conceiving businesses. CRM can be defined according to Gummesson as "CRM is the values and strategies of relationship marketing – with particular emphasis on customer relationships- turned into practical application".

Sterne's definition of CRM is easier than Gummesson, "CRM is the art of using every bit of information that comes into your company about each customer as means of tailoring your communications with them on a one-to-one basis". Today, in the business world, management recognizes that customers are the core of a business and that a company's success depends on effectively managing relationships with them. All objectives are focused to one ultimate goal that is to make customers happy because they are the ones who keep the business running.

Essentially, CRM focuses on building sustainable customer and long-term relationships that add value for both the company and the customer. Despite widespread agreement that CRM can have a direct and indirect impact on customer satisfaction, loyalty, sales and profit, the significance of CRM and the factors that lead to its successful implementation is an area of widespread debate [3].

2. Applications of Data Mining and Trajectory Data Mining in Customer Relationship Management

Moving objects are common in the real world and examples for moving objects are- people, vehicles, animals, goods, ships, materials, and plans etc., The Computational analysis of movement data of movement objects is not well developed research area. Both objects and space details must be considered. Present study discusses about mining trajectories for movement patterns.

- Location based social network services are becoming popular. Understanding human mobility is very useful in location recommendation.

Analyzing movement patterns of vehicles, people, and animals is very useful for understanding their movements, behaviors, and moving paths respectively. Advanced technologies allow finding locations of people or vehicles at any time in many real time applications such as transportation, security, marketing and ecology etc. Initially the goal is to capture the frequent user moving patterns from a set of log data. A movement log is needed for capturing moving patterns of people, animals, vehicles and ships etc.

- Publicly available very large real trajectory data sets are not available
- Nowadays intelligent transportation systems are needed.
- A trajectory is defined as a sequential time-stamped geo-locations in the three dimensional spatio-temporal space.
- In general, trajectory data are stored by using spatio-temporal index such as R-tree, Oct-tree, and Quad-tree etc.

From the CRM point of view, the data mining applications include but not limited to the following:

Customer Retention: Sophisticated customer-retention programs begin with modeling those customers who have defected to identify patterns that led to their defection. These models are then applied to the current customers to identify likely defectors so that preventive actions can be initiated.

Sales and Customer Services: In today's highly competitive environment, superior customer service creates the sales leaders. When information is properly aggregated and delivered to front-line sales and service professionals, customer service is greatly enhanced. If customer information is available, rule-based software can be employed to automatically recommend products. The programs like market-basket analysis have already shown phenomenal gains in cross-selling ratios, floor and shelf layout and product placement improvements and better layout of catalog and web pages.

Marketing: Marketing depends heavily on accurate information to execute retention campaigns, lifetime value analysis, trending targeted promotions, etc. Only by having a complete customer profile can promotions be targeted and targeting dramatically increase response rates and thus decreases campaign cost.

Risk Assessment & Fraud Detection: An accessible customer base significantly reduces the risk of entering into undo risk. For example, a bank can identify fiscally related companies that may be in financial jeopardy before extending a loan to them.

Data Mining Techniques for CRM: Data mining techniques deal with discovery and learning. Data mining techniques may be helpful to accomplish the goal of CRM by extracting or detecting hidden customer characteristics and behaviors from large databases. Some of the popular data mining techniques are :

Association Rule Learning: Association Rule Learning is a popular method for discovering interesting relations between variables in large databases. Agrawal et al. introduced association rules for discovering regularities between products in large scale transaction data recorded by point-of-sale (POS) systems in supermarkets.

Classification and Prediction: Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. It aims at building a model to predict future customer behaviors through classifying databases records into a number of predefined classes based on certain criteria. Classification predicts categorical (unordered) labels, prediction models continuous valued functions. Basic techniques for data classification are decision tree classifier, Bayesian classifier, Bayesian belief networks, rule-based classifiers, fuzzy classification, and support vector machines. Methods for prediction include linear regression, non-linear regression and so on.

Clustering: Clustering is the method by which similar type of records are grouped together. Usually, clustering is done to give the end user a high-level view of what is going on in the database. Clustering is useful for coming up with a birds-eye view of the business.

Regression Analysis: Regression analysis helps us understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. Regression analysis is widely used for prediction and forecasting.

Visualization: Visualization refers to the presentation so that users can view complex patterns. According to Friedman (2008) the main goal of data visualization is to communicate information clearly and effectively through graphical means. It is used with other data mining models

to provide a better and clearer understanding of the discovered patterns or relationships [8].
There are four main categories of objects constituting the majority of trajectory data; 1) human, 2) transportation, 3) animal, and 4) natural phenomena [6]. Human trajectory data is concerning the movements people do as they travel by foot. Transportation trajectory data is also connected to humans moving but is specifically concerning movements made with vehicles. This kind of data could be used in urban planning, resource allocation, traffic analysis, and improving transportation networks. Animal trajectory data is collected by equipping animals with tracking devises and is made for purposes such as studying animal's migratory traces, behavior and living situation. Natural phenomena trajectory data however might need to be collected in alternative ways, indirectly, since it is hard to attach tracking devices natural phenomenon such as hurricanes or tornados. These trajectories are helping scientists deal with nature disasters and protect the natural environment [14].
"CRM is about acquiring and retaining customers, improving customer loyalty, gaining customer insight, and implementing customer-focused strategies. A true customer-centric enterprise helps the business entities drive new growth, maintain competitive agility, and attain operational excellence." Customer Relationship Management (CRM) is a business philosophy involving identifying, understanding and better providing for the customers while building a relationship with each customer to improve customer satisfaction and maximize profits. It's about understanding, anticipating and responding to customers' needs. To manage the relationship with the customer a business needs to collect the right information about its customers and organize that information for proper analysis and decision making. It needs to keep that information up-to-date, make it accessible to employees, and provide the know-how for employees to convert that data into products better matched to customers' needs. Transportation data can be collected, analyzed and the results can be applied to go with better decision making to improve the business.

3. Trajectory Data Clustering

Trajectory data based clustering methodology is very useful for dividing trajectories into groups with similar movement patterns. Discovery of trajectory patterns is very useful in learning interactions between moving objects [26]. Mobility based data clustering is essentially forming the similarity groups of moving objects such as vehicles, animals, people, and cell phones and so on. Many general frameworks exist for mining communities from multiple sources of trajectories. Moving objects are clustered based on trajectory related information such as semantic meaning of trajectories, weights of locations, movement velocity, feature movements, feature characteristics, temporal duration, and spatial dispersion. The mobility-based clustering is less sensitive than the density-based clustering to the size of trajectory dataset [22]. Huey-Ru and Wu et al. [4] proposed an algorithm called Divclust for finding regional typical moving styles by dividing and then clustering the trajectories. Panagiotakis C et al. [13] proposed a method for trajectory segmentation and sampling based on the representativeness of the trajectories in the moving object databases.

3.1 Adopted Methodology for Trajectory Data Clustering

A bus route is a sequence of bus stages. A weight called minimum threshold is assigned for each stage of the each bus route. That is, before clustering all routes of all bus stages, a threshold is selected and is given in the form of weights and in particular weights greater than the specified threshold are considered. These weighted trajectories of bus stages are given as input to the stage clustering process. Bus stages are clustered based on the pre-specified threshold value of each bus stage. In a given set of bus stages all stages whose threshold values are greater than the pre-specified value are clustered into groups. Haozhou Wang et al. [5] proposed algorithms for finding trajectory similarity measures under different circumstances using a common real world taxi cab trajectory dataset.
A new type of trajectory similarity measure is proposed for clustering bus stages. First similarity measure is computed bus routes of trajectories. Each bus route stage details are represented in the form of tree data structure. After creating a set of trees, these trees are clustered based on the similarity measure. For uniform computations and comparisons normalized similarity values are used during the process of clustering. Jun Pang et al. [6] studied the impact from communities on a user's mobility and predict his locations based on his community information.
Q. Li et al. [7] proposed a framework called hierarchical-graph-based similarity measurement (HGSM) for geographic information systems to consistently model each individual's location history and effectively measure the similarity among users.
Wen-Chih Peng et al. [19] proposed an algorithm to capture the frequent user moving patterns from a set of log data in a mobile environment and it is enhanced with the incremental mining capability and is able to discover new moving patterns efficiently without compromising the quality of results obtained. Movement behaviors are characterized using movement sequential patterns and transition probabilities [20].
Zhiwei Lin etc. Al. [30] proposed to represent trees as multidimensional sequences (MDSs), and measure their

similarity based on this representation. Zheng K., et al. [29] proposed an algorithm on representing the uncertainty of the objects moving along road networks as time dependent probability distribution functions. V. Tanuja and Prof. P. Govindarajulu et al. [17] proposed an algorithm for clustering sequential movements of objects such as vehicles, people, and animals and so on.

3.2 Algorithm for Clustering APSRTC Bus Stages

Input: A set of 'n' number of bus routes in the form of trees, each tree represents a root consisting of a set of stages
Output: A set of clustered bus stages

1. Compute occupancy details of each stage for each route from the given sets of stage details.
2. Set the threshold value based on the computed occupancy details
3. Select all the stages of all the routes that satisfy the threshold occupancy value.
4. Create initially 'n' number of individual clusters such that each cluster consisting of a set of related stages.
5. Store details of all the initial clusters of routes in the suitable data structures
6. For each root number i = 1 to n do
7. For each root number j = i + 1 to n do
8. Find similarity measures between roots i and j and store these measures in the appropriate data structure for further processing
9. End-of j-for loop
10. Convert all similarity measures into normalized measures for uniformity
11. Sort all the normalized measures and then select the one with highest similarity measure value for Clustering the two corresponding previous clusters.
12. End-of i -for-loop
13. Repeat the steps 6 through 12 until specified number of final clusters of routes of stages are formed

4. CRM Related Data

In this work, the data pertaining to bus trips of APSRTC, a Road Transport Corporation in Andhra Pradesh, India is collected. APSRTC was formed on 11th January, 1958 as per Road Transport Corporations Act 1950. Earlier, it was a part of Nizam State Rail and Road Transport Department. The organization is divided into four zones with twelve Regional Managers. It has a total of 11,154 scheduled buses operating in 44.15 lakh kilometers and has a total of 426 bus stations and 126 bus depots, entered Guinness Book of World Records on 31 October 1999, with 22,000 buses making it the largest bus fleet in the world [Wikipedia]. MingqiLv et al. [9] proposed route pattern mining framework-which is designed to adapt to the high degree of uncertainty of personal trajectory data, proposed a trajectory abstraction technique-which uses a group-and-partition approach to detect common sub-segments. Destination prediction is similar to the path discovery [11].

Vijay Kovvali et al. [18] listed the data elements and data formats for the vehicle trajectory and supporting data, and provides a structured approach for future video data collection efforts and then the ultimate objective is to provide a research framework for collecting vehicle trajectory and supporting data for behavioral algorithm and safety research. Transportation trajectory data is also connected to humans moving but is specifically concerning movements made with vehicles [27]. A location-based query attempts to find trajectories that are close to all query locations where the query is a small set of locations with or without a specific order constraint and one typical application is route recommendation for a trip to multiple places [28].

4.1 Data sampling

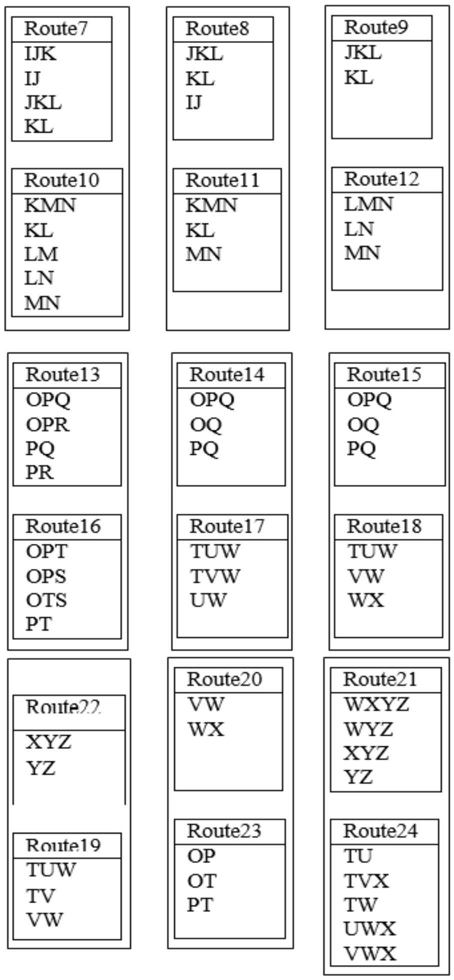
TABLE-1

STAGE	CODE	STAGE	CODE
ALAG	A	INAM	N
ALLU	B	KAKU	O
ALRX	C	KAMM	P
ANNA	D	KCP	Q
ARP	E	KORU	R
BAIT	F	KPB	S
BARM	G	KPD	T
BRAM	H	KPGR	U
CHIN	I	NELL	V
DAND	J	PARL	W
DINN	K	RAJU	X
GANG	L	RAMA	Y
GUND	M	TALA	Z

TABLE-2 Initial Clusters

Route1	Route2	Route3
ABCD	ABCD	BCDE
ACD	ACD	CDE
BD	BD	CEF
BC		DEF

Route4	Route5	Route6
CDE	EFGH	EF
CEF	EGH	EG
DEF	EFH	FH
CF	FH	



All bus stages are coded and coded stages are given in TABLE-1. Each root is represented as a coded tree data structure. Initially each root is considered as one cluster and then initial clusters are grouped into other big clusters based on similarity measures between clusters. For uniform results these similarity measures are normalized and using these normalized values, all the initial clusters are grouped to form larger final clusters.

The data is collected from the NELLORE region of Andhra Pradesh. In this region about 500 buses are being operated daily covering about 30 routes. For every journey the quantities of passengers boarded and altered at each stage are collected. A set of sample bus stages are selected. The collected data is a set of 150 records representing 24 routes with 5 different trip records for each route. There are 26 stages covered through these routes. The dataset is shown in the following table.

TABLE-3

Route	Instance	Sequence
NELL	1	A(45),B(23),C(39),G(42),H(33),K(50)
NELL	2	A(45),B(45),C(33),G(42),H(33),K(50)
NELL	3	A(45),B(40),C(40),G(42),H(33),K(25)
NELL	4	A(45),B(23),C(31),G(42),H(33),K(50)
NELL	5	A(45),B(23),C(39),G(42),H(47),K(50)

4.2 Data preprocessing:

The record for each trip consists of details of number of passengers boarded as well as number of passengers altered at each stage. A trip is considered as profitable when the desired occupancy percentage is reached. A stage of a trip is profitable if the net value of the passengers at the stage met the desired threshold. A weight factor is introduced in the form of threshold .The threshold is applied to get the weighted sequence or trajectory. The weighted trajectory is the input to the trajectory clustering process. In data preprocessing phase each record is processed to get a weighted trajectory.

For example consider the list of the stages for a particular trip as: A(40), B(49), C(51), D(35), E(47), F(52), G(33), H(50) where the letters represent the stage names and the numbers represent the passenger strength at the stage. If we set the threshold as 40 the resultant weighted trajectory is A, B, C, E, F, and H missing D and G as the number is less than the required weight. The set of weighted trajectories for a particular trip forms an input to get a tree of stages in the next phase (Data processing).

The preprocessed data is converted into trajectory data constituting the weighted trajectory sets. The following table represents the trajectory data set, where the weight is fixed at 35.

TABLE-4

Route	Instance	Sequence
NELL	1	A(45),C(39),G(42),K(50)
NELL	2	A(45),B(45),G(42),K(50)
NELL	3	A(45),B(40),C(40),G(42)
NELL	4	A(45),G(42),K(50)
NELL	5	A(45),C(39),G(42),H(47),K(50)

4.3 Data processing

In this phase the sets of weighted trajectories are the inputs to form trees of stages. All these trees are reformed into other sets of stages, each set or tree representing a weighted cluster. The weighted cluster is a set of stages meeting the required occupancy.

The reformed clusters are shown in the TABLE-5.

TABLE-5

Weighted Cluster	Set of stages
Cluster 1	A,X,R,T,Y
Cluster 2	A,D,G,H,J
Cluster 3	A,E,R,T,Y
Cluster 4	A,B,R,S,U,V

5. Results:

Following individual clusters are given as beginning input to the clustering algorithm.

TABLE-6

Route Number	Nodes in the Route
1	A, B, C, D, AB, AC, BC, BD, ABC, ACD, ABCD
2	A, B, D, AB, AC, BD, ABC, ACD, ABCD
3	B, C, D, E, F, BC, CD, CE, DE, BCD, CDE, CEF, BCDE
4	C, D, E, F, CD, CE, CF, DE, CDE, CEF, DEF
5	E, F, G, H, EF, EG, FH, EFG, EFH, EGH, EFGH
6	E, F, H, I, J, EF, EG,FH
7	I, J, K, L, IJ, IL, JK, KL, IJK, JKL
8	I, J, K, L, IJ, JK, KL, JKL
9	J, K, L, JK, KL, JKL
10	K, L, M, N, KM, KL, LM, LN, MN, KMN
11	K, M, N, KM, KL, MN
12	L, M, N, LM, LN, MN, LMN
13	O, P, Q, R, OP, OQ, PQ, PR, OPQ, OR, OQR
14	O, P, Q, T, OP, OQ, PQ, OPQ
15	O, P, Q, OP, OQ, PQ, OPQ
16	O, P, S, OP, OT, PT, OPT, OPS, OTS
17	T, U, V, W, TU, TV, VW, TUW, TVW
18	T, U, V, W, X, TU, VW, WX, TUW
19	T, V, U, TU, TV, VW, TUW
20	V, W, X, VW, WX
21	W, X, Y, Z, WX, WY, XY, YZ, WXY, WYZ, XYZ, WXYZ
22	X, Y, Z, A, XY, YZ, XYZ
23	O, P, T, U, OP, OT, PT
24	T, U, V, W, X, TU, TV, TW, UW, VW, TVX, UWX, VWX

Final Output Clusters are shown in TABLE-7

TABLE-7

Cluster No	Clustered Routes
1	T, TU, TUW, TV, TVW, TVX, TW, U, UW, UWX, V, VW, VWX, W, WX, X
2	K, KL, KM, KMN, L, LM, LMN, LN, M, MN, N
3	O, OP, OPQ, OPS, OPT, OQ, OQR, OR, OT, OTS, P, PQ, PR, PT, Q, R, S, T, U
4	E, EF, EFG, EFGH, EFH, EG, EGH, F, FH, G, H, I, IJ, IJK, IL, J, JK, JKL, K, KL, L
5	A, AB, ABC, ABCD, AC, ACD, B, BC, BCD, BCDE, BD, C, CD, CDE, CE, CEF, CF, D, DE, DEF, E, F, V, VW, W, WX, WXY, WXYZ, WY, WYZ, X, XY, XYZ, Y, YZ, Z,

All the weighted clusters shown in the table are replicas of sets of stages with desired occupancy. Some of the stages of these clusters belong to the existing routes some are not. As all these stages promising the required occupancy, the APSRTC management may take decision to continue the existing routes by providing better customer satisfaction to retain them. There is a scope to form new routes and trips for better customer service and to improve occupancy.

6. Interpretation Relevant to CRM

The result of the clustering process is the sets of the weighted clusters. These clusters suggest the APSRTC management, the better ways to improve the business. A cluster with good occupancy may suggest a new route plan to adopt. At the same time the management can plan the strategies to retain the customers in the profitable routes. Some clusters with lower weights represent low profit routes. In this situation the management may elucidate the reasons behind it by some social or empirical survey, design strategies to attract customers to increase the occupancy and think about the alteration of route maps in that cluster.

7. Conclusion and Future Work:

In this work a novel method is used to cluster the trajectory data sets. The method is compatible to many types of trajectory data sets available in the real world. A road transportation data is preprocessed to get trajectory records using occupancy of passengers as a weighted factor. The weighted trajectories are clustered using the new algorithm. The results of the application are used to suggest the transportation management to improve CRM practices.

Trajectory data mining in general and trajectory data clustering in particular have a large set of prospects to apply in the modern business decision making. The methods applied in this work and the style of the application of the results to CRM will guide the future research in the areas of trajectory data mining and CRM.

8. Acknowledgement:

We owe a debt of gratitude to Sri G.V. Ravi Varma, Regional Manager, APSRTC, Nellore Region, and Mr. V. Srinivasulu Reddy, Retd. Depot Manager, APSRTC, Nellore for their continuous support in collecting the required transportation data.

References

[1] Ali Shahbazi, Student Member, IEEE and James Miller, Member, IEEE “Extended Subtree: A New Similarity Function for Tree Structured Data” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 4, APRIL 2014

[2] Babita Chopra ,VivekBhambri and BalramKrishan, “ Implementation of Data Mining Techniques for Strategic CRM Issues”, VivekBhambri et al, Int. J. Comp. Tech.

Appl., Vol 2 879-883, IJCTA | JULY-AUGUST 2011 , ISSN:2229-6093.

[3] Farnaz Arab, HarihodinSelamat, Suhaimi Ibrahim, and MazdakZamani , " A Survey of Success Factors for CRM", Proceedings of the World Congress on Engineering and Computer Science 2010 Vol II ,WCECS 2010, October 20-22, 2010, San Francisco, USA]

[4] Huey-Ru Wu, Mi-Yen Yeh, and Ming-Syan Chen, "Profiling Moving Objects by Dividing and Clustering Trajectories Spatiotemporally," IEEE Trans. Knowledge and Data Eng., vol. 25, no. 11, pp. 2615–2628, Nov. 2013.

[5] Haozhou Wang1 Han Su, Kai Zheng1 ShaziaSadiq, XiaofangZhou , School of Information Technology and Electrical Engineering, The University of Queensland, "An Effectiveness Study on Trajectory Similarity Measures" Proceedings of the Twenty-Fourth Australasian Database Conference (ADC 2013), Adelaide, Australia

[6] Jun Pang University of Luxembourg jun.pang@uni.lu& Yang Zhang University of Luxembourg FSTC, "Location Prediction: Communities Speak Louder than Friends"

[7] Li Q., Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma, "Mining User Similarity Based on Location History," Proc. 16th ACM SIGSPATIAL Int'l Conf.Advances in Geographic Information Systems(GIS'08),Nov

[8] Md. Rashid Farooqi and Khalid Raza, "A Comprehensive Study of CRM through Data Mining Techniques",Proceedings of the National Conference; NCCIST-2011,, September 09, 2011 Apeejay School of Management, Sector-8, Dwarka, New Delhi.

[9] MINGQI LV, YINGLONG LI, ZHENMING YUAN AND QIHUI WANG,College of Computer Science and Technology,Zhejiang University of Technology,Zhejiang, 310023 P.R. China, "Route Pattern Mining From Personal Trajectory Data", JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 31, 147-164 (2015)

[10] Monika Urbanowicz, "Concept of Customer Relationship Management as an example of innovation in banking sector", IET Working Papers Series ,No. WPS06/2008, ISBN: 1646-8929].

[11] Noulas. A., S. Scellato, N. Lathia, and C. Mascolo, ``Mining user mobility features for next place prediction in location-based services," in Proc.12th IEEE Int. Conf. Data Mining (ICDM), City of Brussels, Belgium, Dec. 2012, pp. 1038_1043.

[12] Oznur Kirmemis Alkan and Pinar Karagoz, "CRoM and HuspExt: Improving Efficiency of High Utility Sequential Pattern Extraction" , IEEE Trans. Knowledge and Data Eng., vol. 27, no. 10, pp. 2645–2657, October. 2015.

[13] Panagiotakis C., N. Pelekis, I. Kopanakis, E. Ramasso, and Y. Theodoridis, "Segmentation and sampling of moving object trajectories based on representativeness", IEEE Trans. Knowl. Data Eng., vol. 24, no. 7, pp.1328_1343, Jul. 2012.

[14] Petter Kihlstroom, "Literature Study and Assessment of Trajectory Data Mining Tools" Degree Project in Built Environment, First Cycle Stockholm 2015]

[15] SusantaSatpathy, Lokesh Sharma, Ajaya K. Akasapu, Netreshwari Sharma, "Towards Mining Approaches for Trajectory Data", International Journal of Advances in Science and Technology Vol. 2, No.3, 2011]

[16] Tanuja.V and Prof. P. Govindarajulu, " A Survey on Trajectory Data Mining", International Journal of Computer Science and Security, Vol.10, Issue 5,2016

[17] Tanuja. V and Prof. P. Govindarajulu, "Application of Trajectory Data Mining Techniques in CRM using Movement Based Community Clustering ", International Journal of Computer Science and Network Security, Vol. 16 , N0.11, November'2016.

[18] Vijay Kovvali, Ph.D., T.E., VassiliAlexiadis, Ph.D., and Lin Zhang, Ph.D.,P.E., "Video-Based Vehicle Trajectory Data Collection" , Cambridge Systematics, Inc., Kovvali, Alexiadis and Zhang 2

[19] W.C. Peng and M.-S. Chen, "Developing Data Allocation Schemes by Incremental Mining of User Moving Patterns in a Mobile Computing System," IEEE Trans. Knowledge and Data Eng., vol. 15, no. 1, pp. 70-85, Jan./Feb. 2003.

[20] Wen-Yuan Zhu, Wen- Chih Peng, Member, IEEE, Chih-Chieh Hung, Po-Ruey Lei, and Ling-Jyh Chen, Senior Member, IEEE, "Exploring Sequential Probability Tree for Movement- Based Community Discovery", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 11, NOVEMBER 2014.

[21] Xin Cao, G. Cong, and C.S. Jensen, "Mining Significant Semantic Locations from GPS Data", Proc. VLDB Endowment,vol.3, no.1, pp. 1009-1020, Sept. 2010.

[22] YU ZHENG. "Trajectory Data Mining: An Overview" Microsoft Research, ACM Transactions on Intelligent Systems and Technology, Vol 6, No.3, Article 29, May 2015.

[23] Zaiben Chen, Heng Tao Shen, Xiaofang Zhou, Yu Zheng, Xing Xie, "Searching Trajectories by Locations – An Efficiency Study", SIGMOD'10, June 6–11, 2010, Indianapolis, Indiana, USA.

[24] Zheng, S. Shang, N. J. Yuan, and Y. Yang, ``Towards efficient search for activity trajectories," in Proc. 29th IEEE Int. Conf. Data Eng. (ICDE), Brisbane, QLD, Australia, Apr. 2013, pp. 230_241.

[25] Zheng Y., L. Zhang, Z. Ma, X. Xie, and W.-Y. Ma, "Recommending Friends and Locations Based on Individual Location History," ACM Trans. Web, vol. 5, no. 1, article 5, Feb. 2011.

[26] Zheng K., Y. Zheng, N. J. Yuan, S. Shang, and X. Zhou, "Online discovery of gathering patterns over trajectories" , IEEE Trans. Knowl.Data Eng.,vol.26, no. 8, pp. 1974_1988, Aug. 2014.

[27] Zheng Y. 2015. "Trajectory Data Mining: An Overview", ACM Transactions on Intelligent Systems and Technology (TIST) - Survey Paper, Regular Papers and Special Section on Participatory Sensing and Crowd Intelligence, Volume 6 Issue 3, May 2015 Article No. 29.

[28] ZHENNI, FENG, AND YANMIN ZHU (Member IEEE), Department of Computer Sceince snd Engineering, Shanghai Jiao Tong University, "A Survey on Trajectory Data Mining: Techniques and Applications", Volume 4, April, 2016, Digital Object Identifier 10.1109/ACCESS.2016.2553681.

[29] Zhixian Yan, supervised by Prof. Stefano Spaccapietra, EPFL Swiss, Federal Institute of Technology, &Lausanne, Switzerland, "Towards Semantic Trajectory Data Analysis:

A Conceptual and Computational Approach", VLDB 09, August-24-28, 2009, Lyon, France, ACM
[30] Zhiwei Lin, Hui Wang, and Sally McClean, "A Multidimensional Sequence Approach to Measuring Tree Similarity", IEEE Trans. Knowledge and Data Eng., vol. 24, no. 2, pp. 197–208, Feb. 2012.

V. TANUJA received Master of Computer Applications degree from Sri Venkateswara University, Tirupati, AP and Master of Technology degree in Computer Science & Engineering from Acharya Nagarjuna University, Guntur, and AP. She is a research scholar in the department of Computer Science, Sri Venkateswara University, Tirupati, AP, India. Her research focus is on Data Mining in Customer Relationship Management. .

P. GOVINDARAJULU, Professor, Department of Computer Science, Sri Venkateswara University, Tirupathi, AP, India. He received his M. Tech., from IIT Madras (Chennai), Ph. D from IIT Bombay (Mumbai), His area of research are Databases, Data Mining, Image processing, Intelligent Systems and Software Engineering.